

Международный контент

УДК 026.06(100)

ББК 78.347.8(0)

DOI 10.25281/0869-608X-2019-68-1-67-76

Т.Е. САВИЦКАЯ

Проект «Миллион книг» — яркий пример международной кооперации

Реферат. В статье анализируются концепция, этапы становления, организационные и технологические особенности проекта «Миллион книг» (Million Book Project). Это оригинальная версия мультязычной полнотекстовой электронной библиотеки, первый этап более обширной программы Универсальной цифровой библиотеки (Universal Digital Library). Целью проекта было оцифровать к 2007 г. 1 млн книг и обеспечить к ним свободный доступ. Он интересен, во-первых, как ранний продуктивный опыт создания компактной электронной библиотеки для оптимизации учебного процесса, во-вторых, как пример широкой действенной международной кооперации в организации информационных ресурсов. В основу коллекции были положены «Книги для библиотек колледжей» (Books for College Libraries) — специальная подборка учебной научной литературы (50 тыс. названий), электронные копии которой были предоставлены некоммерческим членским компьютерным сервисом OCLC (Online Computer Library Center). Проект «Миллион книг» рассматривается как оригинальная модель международного разделения труда в сфере формирования библиотечных информационных ресурсов, когда центры сканирования в Индии и Китае взяли на себя большую часть работы по оцифровке книг, в том числе и из США, а Университет Карнеги — Меллона в сотрудничестве с другими американскими университетами предоставил им необходимую аппаратуру и обучил персонал. Проект основан на изначальном полномочном участии, наряду с библиотекой Университета Карнеги — Меллона и библиотеками других американских университетов, цифровых собраний Индии, Китая, Египта. Одна из сильных сторон проекта — сотрудничество с Продовольственной и сельскохозяйственной организацией ООН, нацеленной на борьбу с голодом, а также с Национальной сельскохозяйственной библиотекой США в оцифровке специальной литературы по проблемам модернизации практики сельского хозяйства, лесоводства и рыбоводства. Значимость проекта обусловлена также тем, что в сфере организационных и технологических решений, связанных с формированием электронной библиотеки, он послужил тестовой базой прикладных исследований в области улучшения техник сканирования, оптического распознавания символов и машинного перевода. Важное преимущество проекта «Миллион книг» — предельная децентрализация полномочий в сфере комплектования коллекции, в результате в его орбиту попадают уникальные памятники истории и культуры стран Востока. Оцифровка и включение в электронную библиотеку культурных сокровищ впервые делает их потенциально доступными глобальной аудитории.



Татьяна Евгеньевна Савицкая,
Российская государственная библиотека,
Центр по исследованию проблем развития библиотек в информационном обществе,
ведущий научный сотрудник
Воздвиженка ул., д. 3/5,
Москва, 119019, Россия
ORCID 0000-0001-5357-2182
E-mail: eneklessa@yandex.ru

Ключевые слова: электронные библиотеки, оцифровка, машинный перевод, цифровая коллекция, международное сотрудничество, информатизация, электронные ресурсы, модернизация, мультязычная полнотекстовая библиотека, Университет Карнеги — Меллона, библиотеки образовательных учреждений, книжные памятники.

Для цитирования: Савицкая Т.Е. Проект «Миллион книг» — яркий пример международной кооперации // Библиотековедение. 2019. Т. 68, № 1. С. 67—76. DOI: 10.25281/0869-608X-2019-68-1-67-76.

Информационная парадигма электронных библиотек в США была бы неполной без оригинального проекта «Миллион книг» (Million Book Project), запущенного еще в 2001 г. Школой компьютерных наук Университета Карнеги — Меллона (Питтсбург, штат Пенсильвания, США). Проект, целью которого было оцифровать к 2007 г. 1 млн книг и обеспечить к ним свободный доступ, интересен, во-первых, как ранний продуктивный опыт создания компактной электронной библиотеки в целях оптимизации учебного процесса, во-вторых, как пример широкой действенной международной кооперации в организации информационных ресурсов.

В отличие от мультязычного проекта Европеана, где электронные ресурсы различных стран объединены лишь на уровне метаданных, а также Всемирной цифровой библиотеки под эгидой Библиотеки Конгресса США и ЮНЕСКО, нацеленной на электронную репрезентацию лишь вершинных произведений культуры народов мира, проект «Миллион книг» базируется на изначальном полноправном участии, наряду с библиотекой Университета Карнеги — Меллона и других американских университетов, цифровых собраний Индии, Китая, Египта. Учитывая, что Европеана открылась в 2008 г., а доступ к сайту Всемирной цифровой библиотеки стал возможен не ранее 2009 г., проект «Миллион книг» следует признать первым опытом продуктивного международного сотрудничества на ниве информатизации библиотечных ресурсов, при этом важно помнить, что он представляет собой лишь первый этап планируемой Универсальной цифровой библиотеки (Universal Digital Library) — полнотекстовой онлайн-библиотеки в 10 млн книг (<http://www.ulib.org/>) [1].

Концепция и этапы становления

К моменту старта проекта «Миллион книг» в 2001 г. в США еще не существовало развитой

инфраструктуры электронных библиотечных ресурсов. Проект «Гутенберг», первопроходец в области развития электронных библиотек [2], накануне подключения к нему волонтерского движения «Распределенные корректоры» (Distributed Proofreaders) едва достиг отметки в 4 тыс. оцифрованных книг. Работающая с 1996 г. электронная библиотека Internet Archive специализировалась главным образом на сканировании веб-страниц с помощью бота (знаменитой Wayback Machine), не сформировав еще пакет таких уникальных проектов, как Archive-It и Open Library. Библиотека Конгресса США разрабатывала модель двуязычных цифровых коллекций по образцу российско-американского проекта «Встреча на границах», сайт которого был открыт в декабре 1999 года. До впечатляющего старта Google Book Search, амбициозного книжного проекта мегакорпорации Google, оставалось целых три года (подробнее об этом см.: [3; 4]).

Именно в этом контексте следует оценивать инициативу ведущих сотрудников Школы компьютерных наук Университета Карнеги — Меллона продолжить благородную инициативу Проекта «Гутенберг» [5], сформировав полнотекстовую общедоступную электронную библиотеку заранее лимитированной информационной емкости. Что подтолкнуло команду единомышленников к такому решению? Как отмечает Д. Трол Кови, главный библиотекарь по специальным проектам университета, проведенные в начале 2000-х гг. социологические исследования выявили, что 75% студентов в процессе обучения чаще используют Интернет, чем университетские библиотеки, при этом 96% из них считают полученную информацию достаточной [6].

Специалисты в сфере образовательных услуг пришли к выводу, что лишь 6% контента поверхностного Интернета (the surface web), индексируемого популярными поисковыми машинами, пригодно для научной работы студентов. Преподаватели убеждены, что недостаток качественных информационных ресурсов в

нем оказывает негативное влияние на качество обучения студентов [1]. Создание компактной электронной библиотеки, укомплектованной апробированной научной литературой, и было призвано исправить ситуацию, повысив уровень компетентности учащихся. В более поздней аналитической справке по истории проекта прямо указывается, что нехватка качественной информации в поверхностном Интернете и его воздействие на учебу студентов — главный двигатель проекта «Миллион книг» [1]. Именно поэтому в основу коллекции были положены «Книги для библиотек колледжей» (Books for College Libraries), специальная подборка учебной научной литературы (50 тыс. названий), электронные копии книг которой были предоставлены некоммерческим членским компьютерным библиотечным сервисом OCLC (Online Computer Library Center) [7].

Основанный еще в 1967 г. как Библиотечный центр колледжа в Огайо (Ohio College Library Center), OCLC, сохранив аббревиатуру, с наступлением цифровой эры существенно модернизировал и расширил свою деятельность, объединяя в настоящее время более 72 тыс. библиотек, архивов и музеев в 170 странах мира. Совокупными усилиями OCLC разрабатывает и поддерживает WorldCat — крупнейший в мире объединенный электронный каталог, доступный для членов ассоциации. Без поддержки OCLC, предоставляющего библиографическую, реферативную и полнотекстовую информацию, многие начинания в области реновации библиотечной системы в США не были бы столь успешными. Будучи партнером проекта «Миллион книг», OCLC ведет цифровую регистрацию, которая регулярно учитывается в проекте, что помогает предотвратить дублирование оцифровки книг. OCLC также осуществляет постоянное архивирование коллекции проекта «Миллион книг» и обеспечивает доступ к книгам посредством ссылок в WorldCat [1].

Задуманный как прикладной образовательный ресурс, созданный в помощь учащимся, проект «Миллион книг» опирался все же на более широкую идейную основу — концепцию универсальной библиотеки. Этот давний просветительский идеал четко сформулировал в Билле о правах для информационного общества Дж. Карбонел, директор Института лингвистических технологий в Университете Карнеги — Меллона, как предоставление надлежащей (right) информации надлежащим людям в надлежащее время на надлежащем языке в надлежащем объеме [8]. Как отметил Р. Рэди, профессор компьютерных наук и робототех-

ники в Университете Карнеги — Меллона, возглавляющий проект «Миллион книг», он приближает нас к идеалу универсальной библиотеки: сделать все опубликованные работы доступными для любого человека, в любое время, на любом языке; исчезают экономические барьеры распространения знания [9].

Предварительное тестирование проекта в рамках формирования электронных библиотек малого объема («Сто книг» и «Тысяча книг») и трудности, выявленные в процессе их реализации, убедили его учредителей в необходимости широкой международной кооперации ради достижения благородной цели создания универсальной цифровой библиотеки открытого доступа, объединяющей Запад и Восток. В результате в процессе формирования новой электронной библиотеки сложилась оригинальная модель международного разделения труда, когда центры сканирования в Индии и Китае взяли на себя большую часть работы по оцифровке книг, в том числе и из США (доставляемых в эти страны в специальных контейнерах по морю), а Университет Карнеги — Меллона при поддержке других американских университетов представил им необходимую аппаратуру и обучил персонал. При этом китайские партнеры оцифровывали также редкие и уникальные коллекции из старейших библиотек, а партнеры из Индии — государственные документы на 11 из 18 официальных языков этой страны.

Проект получил достаточное финансирование: Фонд национальной науки США предоставил 3,63 млн долл. США на закупку оборудования и транспортные расходы по перемещению книг из США в страны Азии и обратно. Индия выделила 25 млн на четыре года на поддержку перевода научно-исследовательской и технологической документации по проекту на местные языки, Министерство образования Китая — 8,46 млн долл. США сроком на три года. Internet Archive предоставил оборудование, персонал и денежные средства. Калифорнийский университет в Мерседе взял на себя финансирование работы по прояснению правового статуса книг и достижению договоренностей с правообладателями.

Таким образом, партнерами в создании мультязычной полнотекстовой электронной библиотеки выступили:

- со стороны Китая — Министерство образования, Академия наук, университеты Фуданьский, Нанкинский, Пекинский, Цинхуа и Чжецзянский;
- со стороны Индии — Индийский институт науки (Бангалор), Международный ин-

ститут информационных технологий, Индийский институт информационных технологий, Университет Анны (г. Ченнаи), университеты Пуны, Гоа, Майсурский, Шаньмуская академия искусств, наук и исследований, Колледж инженерии Арулмигу Каласалингам, Корпорация промышленного развития Махараштры, Храмовый комплекс Тирумалы Венкатешвары;

- со стороны США (помимо Университета Карнеги — Меллона, осуществляющего руководство проектом) — университеты Индианы, штата Пенсильвания, Стэнфордский, Калифорнийский в Беркли, Калифорнийский в Мерседе, Питтсбурга, Вашингтонский, Консорциум «Тройственный колледж» (объединение трех частных колледжей искусств в пригороде Филадельфии);

- со стороны Европы — университеты Копенгагенский, Орхусский, Оденсе (Дания), Датская виртуальная библиотека (<http://www.vlibRARY.net>) [1].

Особо значимый партнер, присоединившийся позднее, после Всемирного саммита по информационным наукам (World Summit on the Information Science, WSIS), состоявшегося в ноябре 2005 г. в Тунисе — Александрийская библиотека (Египет), участие которой, помимо знаменитого бренда, наполнило проект редкими книгами на арабском языке.

Объединение столь масштабных усилий быстро принесло плоды. К ноябрю 2005 г., когда в Китае действовало 18 центров по сканированию книг, а в Индии — 22 (сканирование выполнялось также в Египте, на Гавайских островах и в Университете Карнеги — Меллона), было оцифровано более 600 тыс. книг: 170 тыс. в Индии, 420 тыс. в Китае и 20 тыс. в Египте. Из них около 135 тыс. книг на английском языке, остальные — на языках народов Индии, китайском, арабском, французском и иных языках [8]. Собранная коллекция отображалась как минимум на четырех сайтах-зеркала: в Индии, Китае, Университете Карнеги — Меллона и Internet Archive. Большую часть собрания составляла общественно доступная литература, однако было получено разрешение от правообладателей на включение в коллекцию свыше 60 тыс. книг, имеющих знак охраны авторского права. Из них 53 тыс. книг были на английском языке, 7 тыс. — на языках народов Индии.

Одна из сильных сторон проекта — сотрудничество с Продовольственной и сельскохозяйственной организацией ООН (United Nations Food and Agriculture Organization, FAO), нацеленной на борьбу с голодом, а также с Национальной сельскохозяйственной библиотекой США в оцифровке специальной литературы по

проблемам модернизации практики сельского хозяйства, лесоводства и рыбоводства. Выкладывая в сеть подобные материалы, доступные для сельских общин развивающихся стран, проект вносит вклад в обеспечение продовольственной безопасности, охрану генетических ресурсов и биоразнообразия.

Динамичный рост темпов оцифровки легко можно выявить, проанализировав данные на 16 декабря 2004 г., приводимые Internet Archive. Тогда индийскими сканирующими центрами было оцифровано 10 727 книг (в основном, гуманитарной тематики), из них на английском — 4 459 книг, на языке телугу — 2 244, урду — 855, санскрите — 281, тамильском — 122, древнем дравидийском языке каннада — 72, хинди — 51, французском — 8, немецком — 10, индонезийском — 7, малайском — 6, арабском — 2, итальянском — 2, яванском — 2 книги и пр. [10]. Если проанализировать лингвистическую структуру коллекции, заметно пропорциональное представительство источников на языках народов Индии при выраженном преобладании литературы на языке телугу; англоязычная литература, очевидно, по большей части прибыла из США. Если сравнить с приведенными выше данными по той же коллекции на конец 2005 г., то очевидно: за один год, индийская часть коллекции резко увеличила объем (170 тыс. книг).

К декабрю 2007 г., превысив заявленный объем в 1 млн книг, проект «Миллион книг» достиг рубежа в 1,5 млн книг, представленных на 20 языках, из них 970 тыс. — на китайском, 360 тыс. — английском, 50 тыс. — телугу (официальном языке индийских штатов Андхра-Прадеш и Телингана), 40 тыс. — арабском языке [9]. Поистине интернациональный состав собранной коллекции — явление беспрецедентное среди электронных библиотек, поскольку в ней на равных правах сочетаются цифровые материалы разных культур и географических регионов как Запада, так и Востока. Глобальный (но не глобалистский) проект ни в коей мере не затронут европоцентризмом и снобистским невниманием к «менее развитым» культурам, следы чего критики небезосновательно находят в проекте Google Book Search, который знаменит предпочтением англоязычной литературы [11].

Организационные и технологические особенности

Как же удалось в столь короткие сроки выстроить замечательную свободную для чте-

ния электронную библиотеку, всемирно распространённую и с открытой платформой [12]? Как получилось привести к единому знаменателю работу столь различных коллективов, культурный багаж и ценностные установки которых были подчас диаметрально противоположны? Используя общие базовые стандарты при оцифровке литературы и хранении данных, применяя единые средства поиска по разнородным коллекциям, разработчики поставили цель создания компактной электронной библиотеки с единым загрузочным диском. Было решено придерживаться принципа открытого и бесплатного доступа к фондам, отказавшись при этом от каких бы то ни было попыток коммерциализации ресурса, не превращая его, в частности, в площадку для предоставления банерной или контекстной рекламы. Тем не менее для доступа к фондам библиотеки требуется регистрация с указанием фамилии пользователя и его электронного адреса [13].

Значимость проекта «Миллион книг» во многом обусловлена тем, что в сфере организационных и технологических решений, сопряженных с конструированием электронной библиотеки, он стал своего рода тренировочным полигоном для более современных проектов. Вполне закономерно, что, имея дело с оцифровкой литературы на языках народов Индии, а также на китайском, арабском и иных языках, использующих собственные алфавиты и сложные графические системы, проект послужил тестовой базой прикладных исследований в области улучшения техник сканирования, оптического распознавания символов, интеллектуального индексирования, машинного перевода и поиска информации [14]. При этом, как правило, многие организационно-технологические вопросы решались коллегиально. Так, Александрийская библиотека 17—19 ноября 2006 г. организовала 2-ю Международную конференцию по Универсальной цифровой библиотеке — форум библиотечных работников и IT-специалистов для обмена опытом в сфере формирования библиотечных электронных ресурсов, где был учрежден ежегодный семинар для разрешения рабочих вопросов по проекту.

Именно Александрийская библиотека, располагающая более чем 170 тыс. оцифрованных арабских книг, внесла существенный вклад в оптимизацию процессов оцифровки, оптического распознавания символов применительно к арабскому языку и машинной обработки информации. Вследствие этого полный рабочий цикл создания оцифрованной

книги был автоматизирован и интегрирован в информационную систему библиотеки. Разработанная внутренняя система управления документооборотом оцифровки позднее была расширена и преобразована в Репозиторий цифровых архивов (Digital Assets Repository, DAR), предназначенный для сохранения различных типов оцифрованных материалов (слайдов, рукописей, карт, аудио- и видеoinформации) [14].

Особое внимание в Александрийской библиотеке уделяется поиску эффективной высококачественной технологии оптического распознавания символов для массового сканирования книг и рукописей на арабском языке. Активно применяя в процессе работы разработанный в Кувейте автоматический считыватель Sakhr (Sakhr's Automatic Reader), в поиске новых технологических решений в мае 2006 г. библиотека заключила партнерство с компанией NovoDynamics, специализирующейся на производстве программного обеспечения для электронной конвертации изображений и букв в редактируемый текст. Была поставлена задача оптимальной адаптации программы VERUS от NovoDynamics, успешно работающей с языками Среднего Востока (арабский, персидский, фарси, пушту и т. д.), к автоматическому считыванию рукописей и старопечатных книг на арабском языке.

Со значительными сложностями проект «Миллион книг» столкнулся также в процессе применения технологии распознавания символов при оцифровке книг на китайском языке и, особенно, на языках народов Индии. В отличие от английского языка, где число распознаваемых характеристик менее 100, в Индии, где существует около 1,5 тыс. разговорных языков и 17 различных шрифтов, количество таких характеристик исчисляется несколькими сотнями. Кроме того, использование таких алфавитов, как деванагари, подразумевает обязательное применение подчеркивающей линии, объединяющей слова в предложения и абзацы. Выделить отдельные сегменты в таком тексте практически невозможно. Особое качество фонетической системы некоторых индийских языков — соединение гласных и согласных во временно существующие дифтонги, что крайне затрудняет сегментацию речевых единиц. В таких условиях стандартные технологии распознавания символов показывают весьма ограниченную эффективность и требуют коренной доработки.

В конце 2007 г., отмечая очевидные достоинства проекта, Дж. Беверли писала, что

несмотря на то, что компании Google, Microsoft и Internet Archive инициировали более значительные проекты по оцифровке книг, «Миллион книг» представляет собой крупнейшую в мире электронную библиотеку из книг, находящихся в свободном доступе, базирующуюся на университетских коллекциях [13]. Однако одна из старейших электронных библиотек лишилась этой преференции: в октябре 2008 г. стартовал проект Hathitrust Library — консорциум 13 цифровых библиотек университетов Среднего Запада США и библиотек 11 университетов штата Калифорния; а в апреле 2013 г. первых пользователей приняла Цифровая публичная библиотека Америки. Безусловно, один из первых опытов построения мультязычной общедоступной электронной библиотеки был в полной мере востребован разработчиками современных моделей информатизации библиотечных ресурсов.

Так, известно, что представители IT-гиганта Google, готовясь к развертыванию собственного книжного сервиса, внимательно ознакомились с концепцией и принципами организации проекта «Миллион книг». Повествуя об истории развития Google Book Search, собственного корпоративного электронного книгохранилища со встроенной функцией дифференцированного поиска, его разработчики отмечают, что, заинтересовавшись необычными проектами оцифровки во всем мире, например, проектом Библиотеки Конгресса США «Память Америки», проектами «Гутенберг», «Миллион книг» и Универсальная цифровая библиотека, команда специалистов стала посещать эти проекты, чтобы больше узнать о том, как они работают [15] (о проекте Библиотеки Конгресса США «Память Америки» и последовавшем за ним российско-американском проекте «Встреча на границах» см.: [16]).

Тем не менее проект «Миллион книг» нельзя считать полностью завершенным. На 3-й ежегодной Международной конференции по Всемирной цифровой библиотеке, состоявшейся в Университете Карнеги — Меллона 2—4 ноября 2007 г., было принято решение его продолжать [13]. Как утверждает в аналитической справке, подготовленной в Университете Карнеги — Меллона, изначальная цель проекта «Миллион книг» достигнута, хотя осталось совершить еще много работы [1]. Партнеры проекта планируют продолжить совместные труды по следующим направлениям:

- интеллектуальная собственность;
- машинный перевод и суммирование контента;

- обеспечение централизованного доступа к метаданным;

- удобство использования;
- рост коллекции;
- поддержка культурного разнообразия.

Интеллектуальная собственность. В США по закону о защите авторских прав материалы, опубликованные после 1963 г., охраняются в течение жизни автора и еще 70 лет. То, что вышло до 1923 г., находится вне зоны защиты авторского права; в 1923—1963 гг. — подлежит обновлению, что на практике в 90% случаев не осуществляется. Кропотливый и дорогостоящий труд по прояснению статуса произведений и возможному получению в случае обнаружения правообладателя разрешения на включение электронного текста в онлайн-библиотеку начал автоматизироваться с помощью программного обеспечения, разрабатываемого Internet Archive, партнером проекта. Для сравнения: в Китае по закону все вновь опубликованные материалы сканируются и распределяются по библиотекам страны. В Индии существует закон, разрешающий сканирование книг, не находящихся в продаже, с выплатой суммы компенсации из специально созданного фонда.

Машинный перевод и суммирование контента. Концепция Универсальной цифровой библиотеки зиждется на беспрепятственном автоматизированном переводе с одного языка на другой как запросов пользователей, так и предоставляемых ответов. В Институте лингвистических технологий Университета Карнеги — Меллона, равно как в ряде университетов Индии и Китая, ведутся углубленные исследования по контекстному машинному переводу и автоматическому суммированию контента, призванные радикально модернизировать процесс обучения и научного поиска.

Обеспечение централизованного доступа к метаданным. Согласно первоначальному плану вся цифровая коллекция проекта должна была размещаться в Университете Карнеги — Меллона при наличии сайтов-зеркал в различных регионах мира. Однако передача файлов превратилась в значительную проблему как по техническим, так и по политическим причинам. Поэтому было решено аккумулировать материал на месте его сканирования, обеспечив централизованный доступ к метаданным, чему, безусловно, должен предшествовать процесс их стандартизации.

Удобство использования. Контент проекта «Миллион книг» хранится в файлах TIFF, один файл — одна страница, что весьма не-

удобно. Директор проекта Р. Реди поставил задачу перевода коллекции в более компактные форматы PDF, HTML, что значительно повысит удобство пользования библиотекой.

Рост коллекции. Участниками проекта планируется дальнейшее пополнение коллекции, особенно в области оцифровки специальной литературы по проблемам модернизации сельского хозяйства, в связи с чем предполагается продолжить активное сотрудничество с Продовольственной и сельскохозяйственной организацией ООН и Национальной сельскохозяйственной библиотекой США.

Поддержка культурного разнообразия. Предполагается и впредь продолжать стратегическую линию на сотрудничество в рамках проекта различных культур Запада и Востока, совершенствуя технологии машинного перевода и суммирования контента для приращения знания и улучшения взаимопонимания.

Дальнейшее развитие проекта «Миллион книг» предполагает также его индексацию популярными поисковыми машинами, в частности Google. Как отмечается в сравнительном исследовании электронных библиотек, вышедшем в свет в 2010 г., проект масштабен по размерам, однако многие его детали неясны, в частности: какое количество отсканированных к настоящему времени книг конвертировано в форму электронного текста, какое внимание уделяется точности и в какой степени данные на выходе корректируются вручную. Указывается, что 10 тыс. из 1 млн книг будут доступны на более чем одном языке, предоставляя текстовое поле для машинного перевода и поиска информации на разных языках [17]. Очевидно, впереди у проекта, новаторского во многих своих начинаниях и замыслах, — большой объем работы по устранению недочетов, достижению большей согласованности между участниками и т. д. Перед командой единомышленников стоят поистине грандиозные задачи. На сайте Универсальной цифровой библиотеки отмечается, что в течение 10 лет ожидается дальнейший рост коллекции вплоть до 10 млн книг. Результатом должен быть уникальный ресурс, доступный для каждого человека в мире 24 часа в сутки 7 дней в неделю, невзирая на национальность или социально-экономические условия [18].

Значительное преимущество проекта «Миллион книг» — предельная децентрализация полномочий в сфере комплектования коллекции, когда каждый партнер определяет параметры своего вклада в контент; объемы того, что должно быть включено, и отбирает материалы, подлежащие сканированию [19]. В результате в орбиту проекта попадают уни-

кальные памятники истории и культуры стран Востока. Так, в состав индийской части коллекции входят памятники на санскрите и тамильском языке из библиотеки Танджура, датированные IV в. до н. э.; древнейшие медицинские тексты Аюрведы на пальмовых листьях; старинные португальские книги из Университета Гоа и др. Оцифровка и включение в электронную библиотеку всех этих культурных сокровищ впервые делает их потенциально доступными глобальной аудитории. Интернациональный проект «Миллион книг» сделал шаг на пути к построению всеобъемлющей всемирно доступной цифровой библиотеки человеческого знания [20].

Список источников

1. Frequently Asked Questions about the Million Book Project [Электронный ресурс] // Carnegie Mellon Libraries. URL: https://libwebspace.library.cmu.edu/libraries-and-collections/MBP_FAQ.html (дата обращения: 06.10.2018).
2. Савицкая Т.Е. Проект «Гутенберг»: старейшая электронная библиотека США // Библиотековедение. 2017. Т. 66, № 5. С. 560—566. DOI: 10.25281/0869-608X-2017-66-5-560-566.
3. Савицкая Т.Е. Проект Google Book Search: за и против // Обсерватория культуры. 2016. Т. 13, № 4. С. 420—428. DOI: 10.25281/2072-3156-2016-13-4-420-428.
4. Савицкая Т.Е. Уроки Google Book Search: что дальше? // Обсерватория культуры. 2016. Т. 13. № 5. С. 534—538. DOI: 10.25281/2072-3156-2016-13-5-534-538.
5. Lebert M. History of Project Gutenberg from 1971—2005 [Электронный ресурс] // Project Gutenberg News. URL: <http://www.gutenbergnews.org/20080202/nef-pg-1971-2005-lebert-en> (дата обращения: 06.10.2018).
6. Troll Covey D. The Million Book Project: Removing Obstacles to Use, Satisfaction and Success [Электронный ресурс] // Slideplayer. URL: <https://slideplayer.com/slide/4970938/#> (дата обращения: 06.10.2018).
7. Reddy R., St. Clair G. The Million Book Digital Library Project [Электронный ресурс] // Carnegie Mellon University. 2001. December 1. URL: <http://www.rr.cs.cmu.edu/mbdl.htm> (дата обращения: 06.10.2018).
8. Encyclopedia. Million Book Project [Электронный ресурс] // Nation Master Encyclopedia. URL: <http://www.statemaster.com/encyclopedia/Million-Book-Project> (дата обращения: 06.10.2018).
9. Online Library Gives Readers Access to 1.5 Million Books [Электронный ресурс] // Carnegie

- Mellon University. URL: https://www.cmu.edu/news/archive/2007/November/nov27_ulib.shtml (дата обращения: 06.10.2018).
10. Million Book Project [Электронный ресурс] // Internet Archive : сайт. URL: <https://archive.org/details/millionbooks> (дата обращения: 06.10.2018).
 11. *St. Clair G.* Million Book Project versus Google Print // Journal of Zhejiang University. Science A: Applied Physics & Engineering. 2005. Vol. 6, № 11. P. 1195—1200. DOI: 10.1007/BF02841655.
 12. *Reddy R., Burns E.* The Million Book Project: The Mini-UL Digital Library Platform [Электронный ресурс] // SlidePlayer : сайт. URL: <https://slideplayer.com/slide/6620112/> (дата обращения: 21.01.2019).
 13. *Beverly J.M.* 1.5 Million Books Digitalized and Still Counting — “The Million Book Project” Hits Its Target and Aims for More [Электронный ресурс] // Bridges. 2007. Vol. 16. December 17. URL: <https://ostaustria.org/bridges-magazine/volume-16-december-17-2007/item/2750-1-5-million-books-digitalized-and-still-counting-the-million-book-project-hits-its-target-and-aims-for-more> (дата обращения: 24.05.2017).
 14. The BA Participates in the World Digital Library and Establishes One for Inscriptions [Электронный ресурс] // Bibliotheca Alexandrina : сайт. URL: <https://www.bibalex.org/en/News/Details?DocumentID=1827&Keywords=> (дата обращения: 06.10.2018).
 15. История поиска книг Google [Электронный ресурс] // Google Книги : сайт. URL: <https://www.google.ru/intl/ru/googlebooks/history.html> (дата обращения: 06.10.2018).
 16. *Савицкая Т.Е.* Всемирная цифровая библиотека: под эгидой Библиотеки Конгресса США и ЮНЕСКО // Библиотекосведение. 2018. Т. 67, № 3. С. 299—307. DOI: 10.25281/0869-608X-2018-67-3-299-307.
 17. *Witten I., Gori M., Numerico T.* Web Dragons: Inside the Myth of Search Engine Technology. Amsterdam ; Boston : M. Kaufman, 2007. 281 p.
 18. UDL. About Us [Электронный ресурс]. URL: www.ulib.org/ULIBAboutUs.htm (дата обращения: 23.09.2018).
 19. *Linke E., Birdie Ch.* Identifying and Selecting Content for the Million Book Project [Электронный ресурс] // Proceedings of the IATUL Conferences 2003. Purdue University : сайт. URL: <https://docs.lib.purdue.edu/cgi/viewcontent.cgi?referer=1&article=1637&context=iatul> (дата обращения: 06.10.2018).
 20. *Crane G.* What Do You Do with a Million Books? [Электронный ресурс] // D-Lib Magazine. 2006. Vol. 12, № 3. URL: <http://www.dlib.org/dlib/march06/crane/03crane.html> (дата обращения: 06.10.2018).

Million Book Project — Vivid Example of International Cooperation

Tatiana E. Savitskaya,

Russian State Library, 3/5 Vozdvizhenka Str., Moscow, 119019, Russia

ORCID 0000-0001-5357-2182

E-mail: eneklessa@yandex.ru

Abstract. The article analyses the conceptual idea, stages of formation, organizational and technological features of the Million Book Project. This is the original version of the multilingual full-text electronic library, the first stage of the more extensive program of Universal Digital Library. The aim of the project was to digitize one million books by 2007 and to provide free access to them. The project is interesting, firstly, as an early productive experience in creating a compact electronic library to optimize the educational process, and secondly, as an example of broad effective international cooperation in organization of information resources. The collection was based on “Books for College Libraries”, which is a special collection of educational scientific literature (50,000 titles), electronic copies of which were provided by the non-profit, membership, computer library service OCLC (Online Computer Library Center). The author considers the Project as an original model of international division of labour in the development of library information resources, when the scanning centres in India and China took over most of the work on digitization of books, including from the United States, and the Carnegie Mellon University, in cooperation with other American Universities, provided them

with the necessary equipment and staff training. The Million Book Project is based on the original full participation, along with the library of Carnegie Mellon University and libraries of other American universities, of digital collections of India, China and Egypt. One of the strengths of the project is cooperation with the United Nations Food and Agriculture Organization (FAO), aimed at combating hunger, as well as with the National Agricultural Library of the United States in digitizing specialized literature on the modernization of agricultural, forestry and fish farming practices. The article also notes the significance of the project in the field of organizational and technological solutions related to the formation of the electronic library: it served as a test base for applied research in the field of improving scanning techniques, optical character recognition and automatic translation. The important advantage of the Million Book Project is the ultimate decentralization of powers in the field of acquisition of the collection, resulted in including into the project orbit of the unique monuments of history and culture of the East. The digitization and inclusion in the electronic library of the cultural treasures for the first time makes them potentially available to the global audience.

Key words: electronic libraries, digitization, automatic translation, digital collection, international cooperation, informatization, electronic resources, modernization, multilingual full-text library, Carnegie Mellon University, libraries of educational institutions, book monuments.

Citation: Savitskaya T.E. Million Book Project — Vivid Example of International Cooperation, *Bibliotekovedenie* [Library and Information Science (Russia)], 2019, vol. 68, no. 1, pp. 67—76. DOI: 10.25281/0869-608X-2019-68-1-67-76.

References

1. Frequently Asked Questions about the Million Book Project, *Carnegie Mellon Libraries*. Available at: https://libweb.space.library.cmu.edu/libraries-and-collections/MBP_FAQ.html (accessed 06.10.2018).
2. Savitskaya T.E. Project Gutenberg: the Oldest Digital Library of the USA. *Bibliotekovedenie* [Library and Information Science (Russia)]. 2017, vol. 66, no. 5, pp. 560—566. (In Russ.). DOI: 10.25281/0869-608X-2017-66-5-560-566.
3. Savitskaya T.E. Google Book Search Project: The Pros and Cons, *Observatoriya kul'tury* [Observatory of Culture], 2016, vol. 13, no. 4, pp. 420—428 (in Russ.). DOI: 10.25281/2072-3156-2016-13-4-420-428.
4. Savitskaya T.E. Lessons of the Google Book Search: What Is Next? *Observatoriya kul'tury* [Observatory of Culture], 2016, vol. 13, no. 5, pp. 534—538 (in Russ.). DOI: 10.25281/2072-3156-2016-13-5-534-538.
5. Lebert M. History of Project Gutenberg from 1971—2005, *Project Gutenberg News*. Available at: <http://www.gutenbergnews.org/20080202/nef-pg-1971-2005-lebert-en> (accessed 06.10.2018).
6. Troll Covey D. The Million Book Project: Removing Obstacles to Use, Satisfaction and Success, *Slideplayer*. Available at: <https://slideplayer.com/slide/4970938/#> (accessed 06.10.2018).
7. Reddy R., St. Clair G. The Million Book Digital Library Project, *Carnegie Mellon University*, 2001, December 1. Available at: <http://www.rr.cs.cmu.edu/mbdl.htm> (accessed 06.10.2018).
8. Encyclopedia. Million Book Project, *Nation Master Encyclopedia*. Available at: <http://www.statemaster.com/encyclopedia/Million-Book-Project> (accessed 06.10.2018).
9. Online Library Gives Readers Access to 1.5 Million Books, *Carnegie Mellon University*. Available at: https://www.cmu.edu/news/archive/2007/November/nov27_ulib.shtml (accessed 06.10.2018).
10. Million Book Project, *Internet Archive*. Available at: <https://archive.org/details/million-books> (accessed 06.10.2018).
11. St. Clair G. Million Book Project versus Google Print, *Journal of Zhejiang University. Science A: Applied Physics & Engineering*, 2005, vol. 6, no. 11, pp. 1195—1200. DOI: 10.1007/BF02841655.
12. Reddy R., Burns E. The Million Book Project: The Mini-UL Digital Library Platform, *SlidePlayer*. Available at: <https://slideplayer.com/slide/6620112/> (accessed 06.10.2018).
13. Beverly J.M. 1.5 Million Books Digitalized and Still Counting — “The Million Book Project” Hits Its Target and Aims for More, *Bridges*, 2007, vol. 16, December 17. Available at: <https://ostaustria.org/bridges-magazine/volume-16-december-17-2007/item/2750-1-5-million-books-digitalized-and-still-counting-the-million-book-project-hits-its-target-and-aims-for-more> (accessed 06.10.2018).
14. The BA Participates in the World Digital Library and Establishes One for Inscriptions, *Bibliotheca Alexandrina*. Available at: <https://>

- www.bibalex.org/en/News/Details?DocumentID=1827&Keywords= (accessed 06.10.2018).
15. Google Book Search History, *Google Knigi* [Google Books]. Available at: <https://www.google.ru/intl/ru/googlebooks/history.html> (accessed 07.05.2017) (in Russ.).
 16. Savitskaya T.E. World Digital Library (WDL): Under the Auspices of the Library of Congress and UNESCO, *Bibliotekovedenie* [Library and Information Science (Russia)], 2018, vol. 67, no. 3, pp. 299—307 (in Russ.). DOI: 10.25281/0869-608X-2018-67-3-299-307.
 17. Witten I., Gori M., Numerico T. *Web Dragons: Inside the Myth of Search Engine Technology*. Amsterdam, Boston, M. Kaufman Publ., 2007, 281 p.
 18. *UDL. About Us*. Available at: www.ulib.org/ULIBAboutUs.htm (accessed 23.09.2018).
 19. Linke E., Birdie Ch. Identifying and Selecting Content for the Million Book Project, *Proceedings of the IATUL Conferences 2003. Purdue University*. Available at: <https://docs.lib.purdue.edu/cgi/viewcontent.cgi?referer=1&article=1637&context=iatul> (accessed 06.10.2018).
 20. Crane G. What Do You Do with a Million Books? *D-Lib Magazine*, 2006, vol. 12, no. 3. Available at: <http://www.dlib.org/dlib/march06/crane/03crane.html> (accessed 06.10.2018).

Анонс



Самарин А.Ю., Фурсенко Л.И. Библиофилы-библиографы конца XIX — начала XX века (П.К. Симони, Д.В. Ульянинский, Н.Н. Орлов) : монография. Москва : МИК, 2019. 156 с.

Книга посвящена трем выдающимся библиофилам-библиографам (П.К. Симони, Д.В. Ульянинскому, Н.Н. Орлову), чья активная деятельность протекала в России в конце XIX — первой трети XX века. В научный оборот вводятся автобиографические и мемуарные материалы, представлены полные библиографии их трудов и публикаций о жизни и творчестве этих замечательных книжников.

Первый раздел посвящен члену-корреспонденту АН СССР Павлу Константиновичу Симони (1859—1939). Среди его научных интересов были: филологическое изучение памятников славянской и древнерусской письменности, публикация и анализ фольклорных произведений, история издательского дела и книжной торговли, палеография, библиография, вспомогательные исторические дисциплины и др. П.К. Симони собрал библиотеку, включавшую около 12 тыс. томов, оставил большое научное наследие, которое состоит как из опубликованных трудов (более 100), так и отложившихся в архиве ученого материалов.

Во втором разделе помещены материалы о Дмитрии Васильевиче Ульянинском (1861—1918). Член-корреспондент АН СССР П.Н. Берков писал о нем: «Особое место в истории русского библиофильства и библиофильской литературы начала XX века занимает Дмитрий Васильевич Ульянинский <...> Его называли “поэтом книги”, его библиотеку — “дворцом книги”, в московских библиофильских кругах он считался главою русских библиофилов начала XX века...».

Третий раздел рассказывает о Николае Николаевиче Орлове (1898—1965), активнейшем участнике библиографической и библиофильской жизни 1920-х — начала 1930-х годов. Круг его интересов был очень широк: экслибрисы и литература о них, деятельность Русского библиографического общества при Московском университете (где он много лет был секретарем) и его история. Н.Н. Орлову принадлежат библиографические указатели современной ему литературы по книжному делу, библиотековедению, библиографии, книжной торговле.

*Издание можно приобрести в книжных магазинах «Москва», «Фаланстер»,
а также в интернет-магазинах и в издательстве.
Справки по телефону: +7 (916) 676-81-04*